SPACE SCIENCES LABORATORY

E73-11163
CR-135757

AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT

(E73-11163) AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT Quarterly Progress Report (California Univ.) 12 p HC $3.00 CSCL 02C
N73-33316
Unclas
G3/13 01163

A report of work done by scientists at the Center for Remote Sensing Research, University of California, Berkeley, under NASA Contract No. NAS 2-7567

Quarterly Progress Report
31 August 1973

UNIVERSITY OF CALIFORNIA, BERKELEY

AGRICULTURAL INTERPRETATION TECHNIQUE DEVELOPMENT
(EPN NO. 382)

Principal Investigator: Robert N. Colwell
Co-investigator: Andrew S. Benson
Contributor: James D. Nichols
Center for Remote Sensing Research
Space Sciences Laboratory
University of California
Berkeley, California 94720

1.0 INTRODUCTION

The data from the Skylab Earth Resources Experimental Package (EREP) is helping to answer questions now posed by many remote sensing specialists and engineers regarding the sensor specifications of future earth resource satellites. For example, using multispectral scanner data from the Earth Resources Technology Satellite (ERTS-1), we can now identify many important agricultural crops with a high degree of accuracy. Therefore, the two obvious questions that arise are (1) Will we be able to identify agricultural crops any more accurately with the EREP S192 multispectral scanner, which nominally has three times the resolving capabilities of the scanner on ERTS-1? and, (2) Will this higher resolution enable us to reduce the number of subsamples required to achieve a particular degree of crop identification accuracy?

Another aspect of the research to be done with Skylab EREP data is that of technique development. Due to the complex nature of the sensor (i.e., the 13 bands from the S192 multispectral scanner and the seven photographic products from the S190A and S190B camera systems), much research must be done to determine what data are the most relevant to monitoring agricultural resources and how that data can be handled most efficiently.

The emphasis of the Skylab agricultural investigations being carried out by the Center for Remote Sensing Research (CRSR) is placed on evaluating quantitatively the Skylab data with respect to its usefulness in agricultural inventories. In addition, we will attempt to determine how well these results compare with those that can be obtained from ERTS-1 data. The following sections of this quarterly progress report explain in detail (1) what has been accomplished to date and (2) those tasks which are projected for the remainder of the contract period.

2.0 WORK PERFORMED DURING THE PERIOD COVERED BY THIS REPORT

2.1 Status of Skylab 2 and Supporting High Altitude Aircraft Imagery

The only Skylab 2 overpass covering our agricultural test sites was Track 63 made on June 3, 1973. At this time, we have received all photographic products from the S190A multispectral camera system and the screening images for the S192 multispectral scanner. The corrected S192 tapes have been ordered, but as of yet have not been received. No S190B imagery was obtained of the test sites.

Aircraft photo missions of the agricultural test sites have been flown to coincide as closely as possible to the overpass of Skylab 2. To date, high quality RC-8 color photographs and 70 mm multiband imagery have been received from the Santa Clara Valley, the Salinas Valley, and the west side of the southern San Joaquin Valley test sites.

2.2 Agricultural Test Sites

Due to the westerly shift of the nominal orbit of Skylab 2, only partial coverage of the two primary test sites was obtained. For the San Joaquin County test site (#101-106), the southwest corner was imaged by the S190A multispectral camera only. For the southern San Joaquin County test site (#101-104), the western half was imaged by the S190A and a narrow strip along the western border was imaged by the S192.

In view of this limited coverage of the primary test sites, the existing sites were expanded, with verbal agreement by personnel at Johnson Spacecraft Center (JSC), to include the agricultural areas in the Santa Clara Valley and the Salinas Valley (see Figure 1). These areas will also be imaged by the Skylab 3 mission thereby providing the sequential coverage that is needed for many of the proposed interpretation tasks.

2.3 Ground Data Collection

Thus far, most of the efforts of the CRSR personnel have been concentrated on collecting ground data in support of the Skylab agricultural experiment. These data are required for the training and evaluation of both human and automatic image analysis. Unfortunately, the data that were collected in the two primary test sites -- San Joaquin County and southern San Joaquin Valley -- prior to the launch of Skylab 1 will be of little use in this experiment due to the unplanned shift in the Skylab 2 ground tracks. However, we were given sufficient notice of this shift to obtain low level aerial photos and to collect ground data in the Santa Clara Valley and the Salinas Valley nearly coincident with the June 2, 1973 Skylab overpass on Track 63.

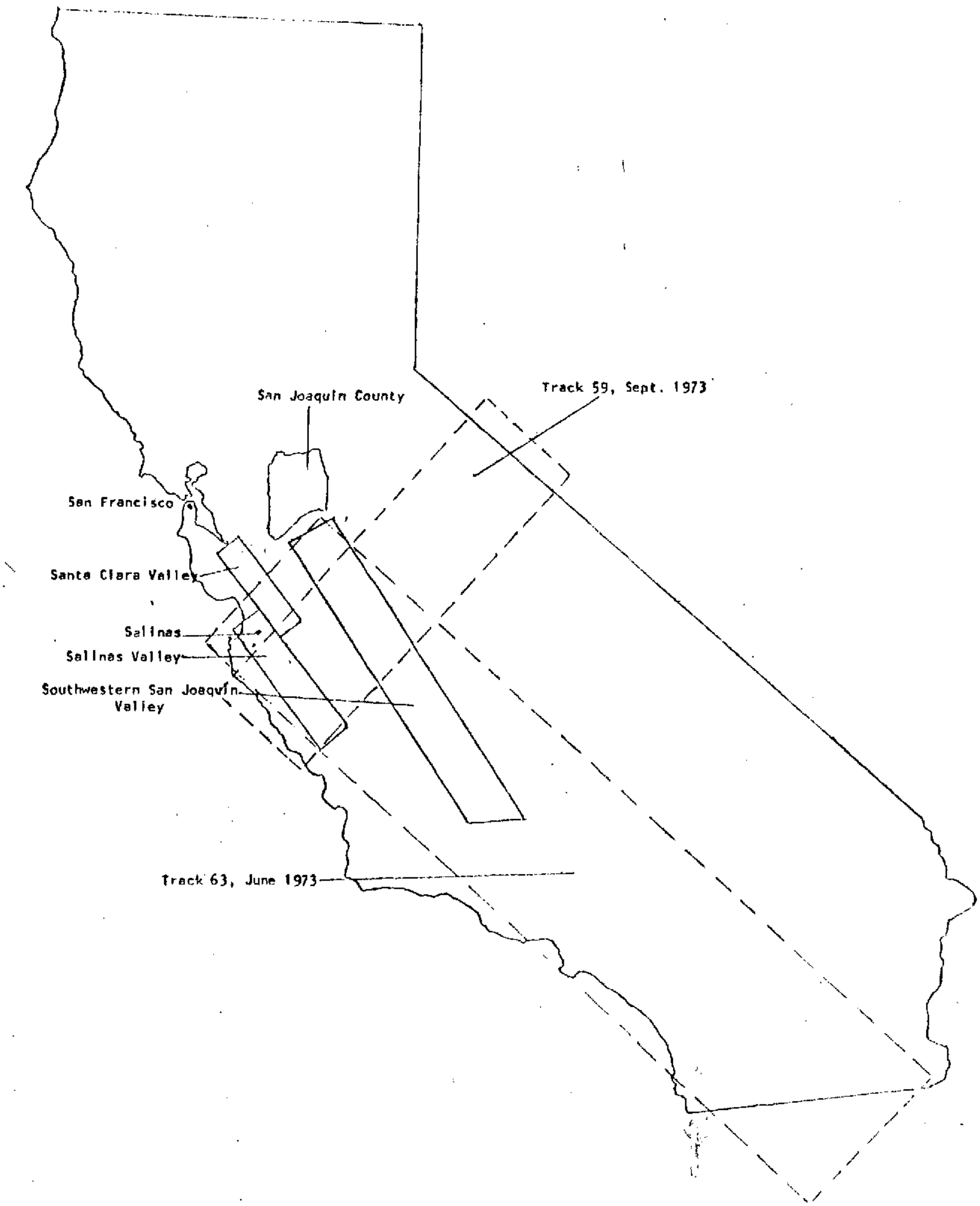

Figure 1. The locations of the new agricultural test sites -- Santa Clara Valley and Salinas Valley -- with respect to Skylab ground tracks. (Solid lines delimit test sites and dashed lines delimit S190A photographic coverage.)

2.4 Personnel Training

An internal training session is being given at the CRSR to image analysts to familiarize them with the CALSCAN* automatic classifier. Prior to this time and when working with ERTS-1 digital tapes, the task of image stratification, which is a prerequisite for efficient computer analysis, was done by human image analysts. The stratified imagery was then given to a computer operator who trained and classified the individual strata with data from the digital tapes. It was felt, however, that these two tasks could be done most efficiently by the image analyst alone because, with his greater knowledge of agricultural cropping practices, he can more effectively train the CALSCAN classifier and make a better evaluation of the final results.

2.5 Automatic Interpretation

The planned automatic data processing studies call for extracting agricultural information from EREP computer-compatible tapes and photographic transparencies. The processing of both types of data will give a comparison of classification results using an analog storage medium (film) versus digital storage (tape).

To date we have not received the digital tapes, but processing of the photographic transparencies has begun. The northern half of the Salinas Valley test site is being used for preliminary evaluation of the technique. This area has been scanned with a microdensitometer on all four bands of S190A photography. The scan interval along the X and Y axes was .01 inches, and the aperture of the scanner was set at .001 inch diameter so that each data point represented a 1.12 acre spot on the ground. The resulting density measurements were recorded on magnetic tape. This tape is now being reformated on the CDC 6600/7600 system so that it will be compatible with the CALSCAN program.

3.0 WORK PLANNED DURING THE NEXT REPORTING PERIOD

3.1 Introduction

Due to the uncertainties that have been associated with the EREP missions, it has been difficult to specify those tasks which will be accomplished throughout the remainder of the contract period. However, now that the areas and time of Skylab coverage are known, we can project fairly accurately those tasks which will be done and their appropriate time frames. A summary of these tasks is shown in Figure 2.

---

*CALSCAN is the CRSR version of the LARS-Purdue pattern recognition program adapted to the CDC 6600/7600 system at the University of California, Berkeley.

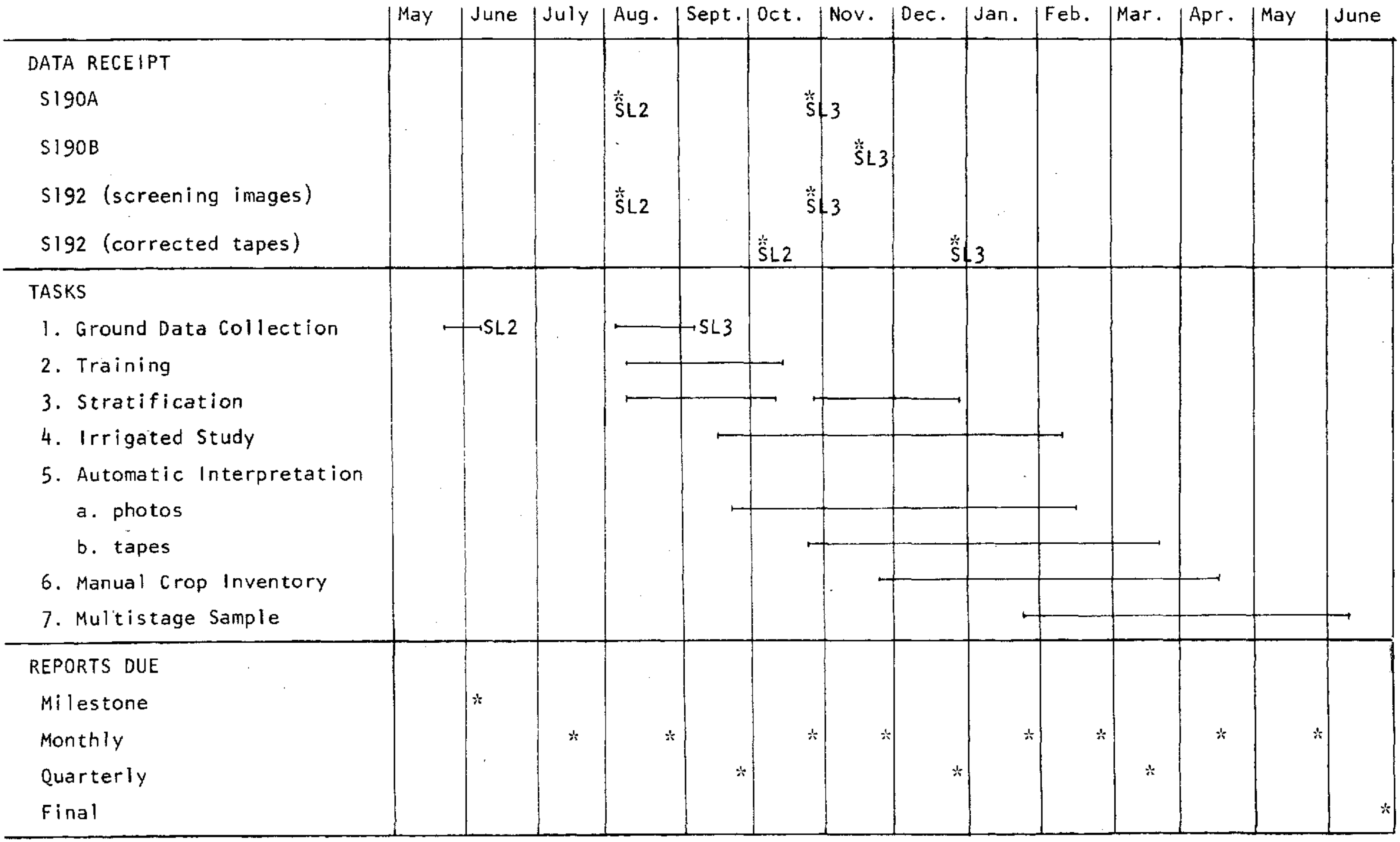

Figure 2. A summary of tasks to be performed.

3.2 Agricultural Land Use Classification

The stratification of agricultural land from other land uses is the first step for all analyses of Skylab data with respect to agricultural resources. Such information is not only necessary for monitoring land use change and the efficient allocation of ground samples, but also is a vital input to the CALSCAN classifier so that resulting statistics can be obtained in the most cost effective manner, without affecting classification accuracy.

The stratification of agricultural lands will be done in all test sites. The bases of these stratifications will be on (1) visual appearance, and (2) the classification system that is now used by the California Crop and Livestock Reporting Service (CLRS). The first stratification system which will be input into the data bank for future use by CALSCAN, will be more detailed than the second. The second stratification is the type that would be used to allocate ground samples for a conventional enumerative survey.

The accuracy of the visual stratification system will be determined indirectly using the CALSCAN classifier. Several different versions of the stratification boundaries, e.g., boundaries drawn by different image analysts, on different dates, on different film types, and on ERTS-1 composites, will be input into CALSCAN to determine how they affect the accuracy of the final output.

The second stratification system will be evaluated by comparing it with the existing stratification boundaries as delineated by the CLRS using MAPIT.

3.3 Irrigated Land Study

There presently exists an important need on the part of the Department of Water Resources of the State of California for a periodic tabulation in any given year of the statewide acreage of agricultural land receiving irrigation. A study will be performed to investigate the extent to which this tabulation can be accomplished using two dates of EREP data and the appropriate sampling designs. This study will not be able to give a valid yearly statistic of the irrigated lands within the test area due to the fact that only two dates of imagery area available and imagery taken at other dates would be required. However, it will help determine (1) the feasibility of making such inventories, (2) the expected accuracies of such inventories, (3) if similar techniques could be used on ERTS-1 imagery, and (4) a projected cost for making statewide inventories.

3.4 Automatic Interpretation

3.4.1 Photographic Data

After the reformatted tapes are available (see section 2.5), the data obtained from the scanned S190A photos will be analyzed with the CALSCAN classifier for purposes of crop identification. If the results of this analysis indicate that the results are comparable to those expected from the S192 multispectral scanner, the remaining agricultural areas not imaged by the scanner but by the photos will be scanned with the microdensitometer.

3.4.2 Multispectral Scanner Data

The initial experimentation with data from the S192 multispectral scanner will be to determine which combinations of the 13 channels are optimum for identifying crop types. This analysis will be made with the CALSCAN classifier.

3.5 Manual Crop Inventory

To date, most of the identification and inventory of agricultural crops using satellite data (i.e., ERTS-1) have been done semi-automatically by computer analysis of digital tapes with human inputs of training materials. Although such systems have achieved very accurate results in certain test areas, there is still a need to develop techniques for manual interpretation of satellite data for agricultural resource inventories. While computer-based systems may ultimately prove to provide the most efficient method for gathering agricultural statistics of extensive areas, at the present time the human interpreter represents the most expedient way to perform an operational inventory in the United States. For many of the emerging nations of the world where both national agricultural statistics and computer systems may be non-existent, the data gathered by human interpreters can provide a valuable input to the management decisions for agricultural resources.

In any attempt to develop efficient techniques for the human interpretation of Skylab data, several factors must be considered: (1) because of the large areal coverage of the imagery, 100 percent image interpretation of the entire frame for detailed information is not practical, (2) a simple method is needed to evaluate the accuracy of the interpreter's estimates and, if necessary, to adjust these estimates, and (3) the low resolution of the imagery makes accurate acreage estimates by human interpreters impossible.

An inventory technique employing a double sampling design utilizing point data is expected to deal effectively with these constraints. For the first stage (large sample) the interpreter will determine the presence or absence of the crop of interest at each of a large number of points throughout the survey area. These data will then be used to estimate the proportion of the area that is planted to that particular crop. This proportion, when multiplied by the total area being inventoried will give an estimate of total crop acreage. The second stage in the inventory will consist of a subsample of the large sample. These subsample points will be field checked, and the correlation between ground conditions and image interpretation estimates will be used to evaluate interpreter accuracy and to calculate a ratio estimator to adjust the interpreter's estimated proportion. It must be noted that, due to the long delay between data acquisition and the receipt of EREP imagery, the second stage data will probably have to be collected from high altitude photography that was obtained coincident with the Skylab overflight.

3.6 Multistage Sampling of Agricultural Resources

In general, there are two types of basic information required by agricultural planners: (1) an estimate of the quantity of specific types of agricultural resources in each administrative unit, and (2) an inplace map of those resources. When a high correlation exists between estimates of the resources as made from spacecraft data and those based on ground observations, unique and valuable information for meeting both of these requirements can be provided in a cost-effective manner.

The culmination of our Skylab agricultural experiments will be the demonstration of agricultural survey techniques using a multistage sampling model. Through use of the discriminant analysis techniques described below, Skylab photo and scanner data will be combined with aerial photos and ground data to give an estimate of crop acreage within the areas common to Track 63 of Skylab 2 and Track 59 of Skylab 3. We have not yet chosen which crop or crops will be inventoried; this will be determined after the data from the S192 multispectral scanner have been analyzed. Under normal operational conditions, the optimum time for obtaining remote sensing data for a successful agricultural inventory is known and is carefully adhered to. However, since this is not possible under the restrictions of the EREP missions, the crops to be surveyed will be those which can best be inventoried with the available Skylab data.

The multistage model relies heavily upon the first stage in which the information extracted from the Skylab data by human interpreters and automatic classifiers provides the initial estimates of the resource. The first step of the data extraction process will be one in which human interpreters stratify all fields within the area of interest into broad land use categories and crop classes based on their appearance on the Skylab S190A ektachrome imagery. At this time political and geographic boundaries will be superimposed on the imagery to further define the areas of interest. Next, a number of fields which represent the various agricultural resources of interest will be selected from each stratum to train the discriminant analysis program. The identities of these fields will be determined from ground data and/or the interpretation of aerial photos. The number of training fields required for each crop class will be dependent upon the variability of the spectral signatures of the crops present. This variability is caused by such factors as different cropping practices, local soil differences, and genetic variations within each particular crop type. For example, a crop such as alfalfa which may be in several stages of maturity throughout the survey area at the time of image acquisition may require five or more fields per stratum for adequate training, whereas only one training field per stratum may be needed for a less complex crop type such as corn. After the fields have been chosen, they will be located on, and extracted from, the spacecraft imagery. The multispectral data from the training fields will be run through the discriminant analysis program to obtain a point-by-point classification of the entire area by strata. This procedure will provide an initial estimate of the acreage of the vegetation classes.

In second stage of the model, the results of the discriminant analysis will be sampled to determine their relationship to ground estimates of the resource. Sampling units (SU) will be defined by dividing the entire area into rectangular blocks. The size and shape of these blocks will be determined by (1) the information requirements, (2) the variability that occurs in the SU estimates as the block size changes, (3) the costs of further subsampling, and (4) the resolution of the scanner imagery.

Probability sampling is expected to provide the most efficient sampling design that can be applied in the second stage of the model. Probability sampling is a special case of the mean of the ratios estimation where samples are allocated proportional to the expected variance of the X_i estimate. For this model, the total value of the i^{th} SU, denoted by X_i, is evaluated by

$$X_i = \sum_{m=1}^{M} \sum_{j=1}^{J} I_m V_j$$

where $\begin{cases} I_m = 1 \text{ if } C_m = j \\ I_m = 0 \text{ otherwise} \end{cases}$

C_m = crop class for the n^{th} "pixel" (picture element) of the SU, as determined by the discriminant analysis,

M = the number of "pixels" per SU,

V_j = the crop class being evaluated

J = the number of crop classes.

The value or weight (v_j) is assigned to rank the various crops or vegetation types based on their relative importance to the survey. In an agricultural inventory where total dollar value is the objective, the v_j's are the average dollar values per "pixel" of the various crops (j). If the making of an inventory of a single crop is the objective, the value of the crop of interest is 1 and all other v_j's are set to zero. In many cases this weighting factor is primarily affected by the marketing conditions of each crop, and is highest for those crops for which acreage estimate errors are most important.

The variance of the population is estimated by

$$S^2 = \frac{1}{N-1} \sum (X_i - \bar{X})^2$$

The number (n) of SU's to be selected for photo and ground measurement when no remote sensing information is available is determined by

$$n = \frac{Nt^2S^2}{N(AE)^2 + t^2S^2}$$

where AE = the allowable error, in units of value

t is a value obtained from "students t" tables and

S^2 is as defined previously.

The n points are then selected from the list of SU's proportional to their estimaged value.

The selected SU's are then carefully transferred to the corresponding high flight photography where precise field size measurements are made for use later in adjusting the acreage estimates obtained from the classifier.

From high flight images, low altitudes images, ground identification and historical data, the "correct" classification for each field in the SU is determined, down to crop type and maturity.

The total value for the area ($\hat{T}$) is estimated using the probability of selection (P_i) and the photo/ground estimate of SU value (Y_i) by means of the relation,

$$\hat{T} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{P_i}$$

where $P_i = \dfrac{x_i}{\sum_{i=1}^{N} x_i}$

The variance of the estimate for $\hat{T}$ is

$$s_{\hat{T}}^2 = \mathrm{Var}(\hat{T}) = \frac{1}{n} \sum_{i=1}^{N} P_i \left(\frac{Y_i}{P_i} - \hat{T}\right)^2$$

If the photo/ground estimate (Y_i) were to be perfectly proportional to the remote sensing estimate (X_i), only one ground sample would be needed to determine the proportionality constant. More realistically however, the number of ground samples (n) for future surveys is estimated by:

$$n = \frac{N t^2 s_{\hat{T}}^2}{N(AE)^2 + t^2 s_{\hat{T}}^2}$$

This probability sampling model is appropriate when a single parameter such as "acreage of a single crop", "value of all the crops present", or "demand for irrigation water" is desired. Such a model can be replaced by a regression sampling model if estimates on a crop-by-crop basis are required; however, the regression sampling model will only meet the allowable error criterion for the total value of all crops by strata.